t-критерии

Проверка гипотез о среднем генеральной совокупности и сравнение двух совокупностей

Критерии для проверки гипотез о среднем

- Сравнение с предполагаемым значением:
  - **z-критерий** (при известном σ)
  - **одновыборочный t-критерий Стьюдента** (one-sample t-test)
- Сравнение средних двух зависимых (связанных) выборок (до/после, родство и т.п.):
  - **t-критерий для зависимых выборок** (paired samples t-test)
- Сравнение средних двух независимых выборок (экспериментальная/контрольная группы, категории):
  - **t-критерий для независимых выборок** (independent sample t-test)

Принцип проверки критериев

1. Сравнение разности наблюдаемого и ожидаемого значений с погрешностью оценки:
$$t = \frac{\bar{X} - \mu_0}{SE}$$
$SE = S/\sqrt{N}$ -стандартная ошибка
2. Оценка с использованием теоретического распределения вероятности получения подобного (или большего) стандартизованного отклонения $t(df), df = N - 1 \rightarrow$ оценка уровня значимости p

Сводная таблица критериев о среднем

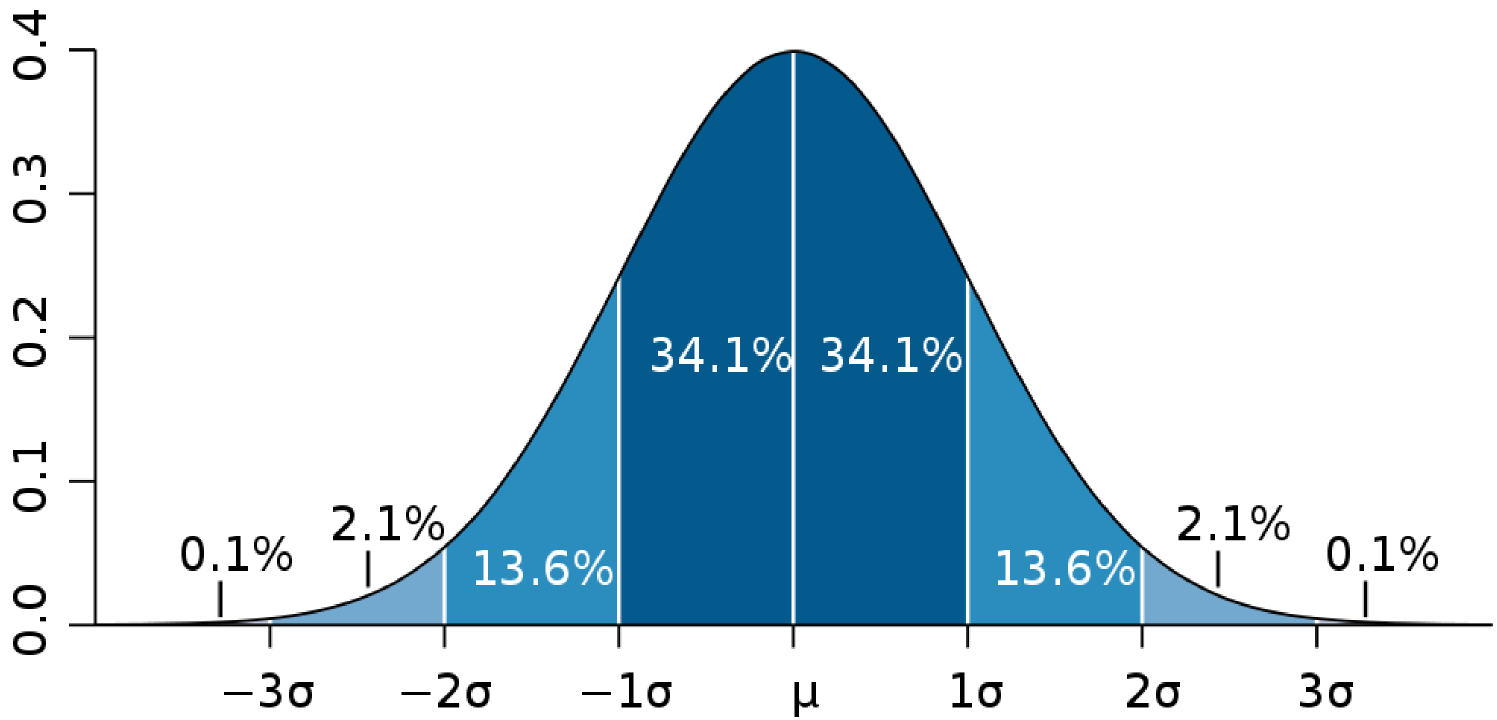
- Во всех критериях используется статистика:

$$t = \frac{\text{наблюдаемое} - \text{предполагаемое}}{\text{стандартная ошибка}}$$

Критерий	Наблюдаемое	Предполагаемое	Ст. ошибка
z	Выборочное среднее	Среднее совокупности	Ст. ошибка для среднего
Одновыборочный t-критерий	Выборочное среднее	Среднее совокупности	Ст. ошибка для среднего
Для зависимых выборок	Средняя разность	Предполагаемая разность (обычно 0)	Ст. ошибка для средней разности
Для независимых выборок	Разность выборочных средних	Предполагаемая разность (обычно 0)	Ст. ошибка для разности средних

Правило трех сигм

- Практически все значения (99,7%) нормальной случайной величины лежат в интервале: $\left(\mu - 3\sigma, \mu + 3\sigma\right)$



- Примерно 95% значений лежит в интервале: $\left(\mu - 2\sigma, \mu + 2\sigma\right)$

Число степеней свободы

Критерий	**df**
z	-
Одновыборочный t-критерий	N-1
Для зависимых выборок	N-1
Для независимых выборок	$(N_1-1) + (N_2-1)$

t-критерий для зависимых выборок

- Такие же формулы, как и для одновыборочного t-критерия
- Наблюдаемые значения – **средние разности** $\overline{D}$

Участник	X_1 (до)*	X_2 (после)	$D = X_2 - X_1$	$(D - \overline{D})$	$(D - \overline{D})^2$
1	19	12	-7	-2	4
2	35	36	1	6	36
3	20	13	-7	-2	4
4	31	24	-7	-2	4
$\boldsymbol{H_0}: \mu_1 = \mu_2, \boldsymbol{H_1}: \mu_1 > \mu_2 \rightarrow$ $\boldsymbol{H_0}: \overline{D} = 0, \boldsymbol{H_1}: \overline{D} < 0$			$\overline{D}$=−5		$SS = 48$

*X_1, X_2 *- число сигарет до и после применения методики для отказа от курения*

$S^2 = \sum(D - \overline{D})^2$/(N-1)=SS/df = 48/3 = 16

$$SE_{\overline{D}} = \frac{S}{\sqrt{N}} = \frac{\sqrt{16}}{\sqrt{4}} = 2$$

$$t = \frac{\overline{D} - 0}{SE_{\overline{D}}} = \frac{-5}{2} = -2.5$$

p-значение для распределения с df=4-1=3 степенями свободы = 0.044
в Excel: СТЬЮДЕНТ.РАСП.2Х(x;df) или СТЬЮДЕНТ.РАСП.ПХ (x;df)

Сравнение одностороннего и двустороннего критериев

- Двусторонний критерий проверяет гипотезу:

  $\boldsymbol{H_0}: \mu = \mu_0$ против $\boldsymbol{H_1}: \mu \neq \mu_0$
- Односторонний критерий проверяет гипотезу:

  $\boldsymbol{H_0}: \mu = \mu_0$ против $\boldsymbol{H_1}: \mu < \mu_0$ или

  $\boldsymbol{H_0}: \mu = \mu_0$ против $\boldsymbol{H_1}: \mu > \mu_0$
- p-значение для одностороннего критерия в 2 раза меньше, чем для двухстороннего, поэтому $\boldsymbol{H_0}$ отвергнуть проще
- Нельзя использовать односторонний критерий, если нет оснований рассматривать только положительные (отрицательные) эффекты изменений. Рекомендуется всегда применять двухсторонний критерий как более строгий
- Решение о типе критерия принимается априори, ДО проведения эксперимента

t-критерий для независимых выборок

- Обозначения:
  $M_1 = \bar{X}_1, M_2 = \bar{X}_2$ - средние двух выборок
- **Разность средних** $\Delta M = M_1 - M_2$
- Статистика критерия:

$$t = \frac{\Delta M - 0}{SE_{\Delta \mathrm{M}}} \sim t(df)$$

Условия	Расчет стандартной ошибки	Число степеней свободы df
$\sigma_1 = \sigma_2$	На основе общей дисперсии S^2	$N_1 + N_2 - 2$
$\sigma_1 \neq \sigma_2$ N_1 и $N_2 > 30$	На основе дисперсий выборок S_1^2, S_2^2	∞ (можно использовать z-распределение)
$\sigma_1 \neq \sigma_2$ N_1 и $N_2 \leq 30$	На основе дисперсий выборок S_1^2, S_2^2	Оценивается по приближенной формуле

Случай 1: $\sigma_1 = \sigma_2$

- Выборки взяты из совокупностей с одинаковой дисперсией σ^2
- В качестве оценки S^2 можно взять средневзвешенное из S_1^2 и S_2^2 ($df_1 = N_1 - 1$ и $df_2 = N_2 - 1$)

$$S^2 = \frac{df_1}{df_1 + df_2} \cdot S_1^2 + \frac{df_2}{df_1 + df_2} \cdot S_2^2$$

- Дисперсии для средних:

$$SE_{M_1}^2 = \frac{S^2}{N_1}$$
$$SE_{M_2}^2 = \frac{S^2}{N_2}$$

- Дисперсия для разности средних:

$$SE_{\Delta M}^2 = SE_{M_2 - M_1}^2 = SE_{M_1}^2 + SE_{M_2}^2$$

Пример (предполагаем $\sigma_1 = \sigma_2$)

- Сравнить время реакции в группах подростков (1)и взрослых людей (2)

Группа	1 (подростки)	2 (взрослые)
Среднее (M)	350 (мс)	360
СКО (S)	20	30
Размер (N)	50	100

- Гипотезы: **H$_0$**: $\mu_1 = \mu_2$, **H$_1$**: $\mu_1 < \mu_2$
- Число степеней свободы: $df_1 = 49, df_2 = 99$

- Общая дисперсия:

$$S^2 = \frac{df_1}{df_1 + df_2} S_1^2 + \frac{df_2}{df_1 + df_2} S_2^2 = 400 \cdot \frac{49}{148} + 900 \cdot \frac{99}{148} \approx 734.5$$

- Стандартные ошибки:

$$SE_{M_1}^2 = \frac{S^2}{N_1} = \frac{734.5}{50} = 14.7 \text{ и } SE_{M_2}^2 = \frac{S^2}{N_2} = 7.3$$

$$SE_{\Delta M} = \sqrt{SE_{M_1}^2 + SE_{M_2}^2} = \sqrt{22} = 4.69$$

- Статистика: $t = \frac{\Delta M}{SE_{\Delta M}} = -\frac{10}{4.69} = -2.13 \rightarrow p \approx 0.017$ (односторонний критерий, df = 148)
- Вывод: отвергаем H_0 в пользу H_1 , различие значимо (не случайно)

Случай 2: $\sigma_1 \neq \sigma_2, N$-большое

- Выборки взяты из совокупностей с разными дисперсиями
- Дисперсии средних необходимо вычислять отдельно, на основе S_1^2 и S_2^2

$$SE_{M_1}^2 = \frac{S_1^2}{N_1}$$
$$SE_{M_2}^2 = \frac{S_2^2}{N_2}$$

- Дисперсия для разности средних:

$$SE_{\Delta M}^2 = SE_{M_2 - M_1}^2 = SE_{M_1}^2 + SE_{M_2}^2$$

- Число степеней свободы - ∞ (z-распределение)

Пример (предполагаем $\sigma_1 \neq \sigma_2$)

- Сравнить время реакции в группах подростков (1)и взрослых людей (2)

Группа	1 (подростки)	2 (взрослые)
Среднее (M)	350 (мс)	360
СКО (S)	20	30
Размер (N)	50	100

- Гипотезы: **H$_0$**: $\mu_1 = \mu_2$, **H$_1$**: $\mu_1 < \mu_2$
- Число степеней свободы: $df_1 = 49 > 30, df_2 = 99 > 30$
- Стандартные ошибки:

$$SE_{M_1}^2 = \frac{S_1^2}{N_1} = \frac{400}{50} = 8$$

$$SE_{M_2}^2 = \frac{S_2^2}{N_2} = \frac{900}{100} = 9$$

$$SE_{\Delta M} = \sqrt{SE_{M_1}^2 + SE_{M_2}^2} = \sqrt{17} \approx 4.12$$

- Статистика: $t = \frac{\Delta M}{SE_{\Delta M}} = -\frac{10}{4.12} = -2.43$
- Уровень значимости: $p \approx 0.008$ (односторонний критерий, z-распределение)
- **Вывод: отвергаем H_0 в пользу H_1, различие значимо (не случайно)**

Случай 3: $\sigma_1 \neq \sigma_2$, N-малое

- Выборки взяты из совокупностей с разными дисперсиями
- Дисперсии средних необходимо вычислять отдельно, на основе S_1^2 и S_2^2
  $SE_{M_1}^2 = \frac{S_1^2}{N_1}$ и $SE_{M_2}^2 = \frac{S_2^2}{N_2}$
  Дисперсия для разности средних:
  $$SE_{\Delta M}^2 = SE_{M_2 - M_1}^2 = SE_{M_1}^2 + SE_{M_2}^2$$
- Для вычисления уровня значимости необходимо использовать t-распределение
- Число степеней свободы (формула Уэлча):
  $$df \approx \frac{\left(SE_{M_1}^2 + SE_{M_2}^2\right)^2}{\frac{\left(SE_{M_1}^2\right)^2}{df_1} + \frac{\left(SE_{M_2}^2\right)^2}{df_2}}$$
- В нашем примере:
  $$df \approx \frac{(4+9)^2}{\frac{4^2}{49} + \frac{9^2}{99}} \approx 148$$

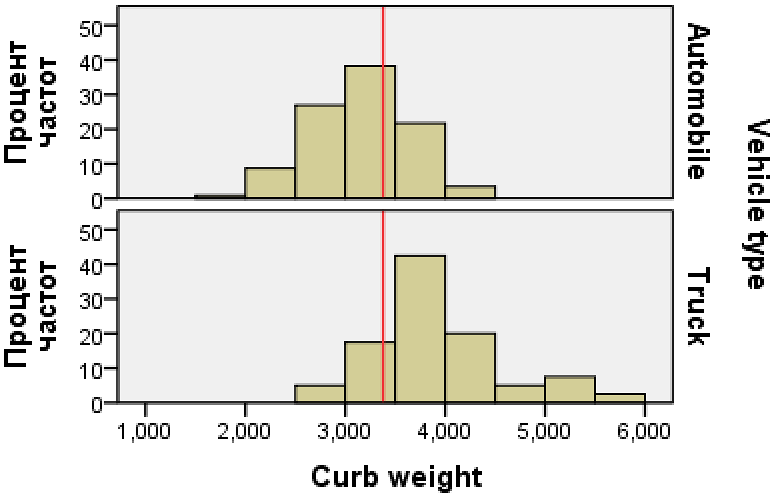
В предположении о равенстве дисперсий было: $df = N_1 + N_2 - 2 = 148$

Равны ли дисперсии совокупностей?

- При различии дисперсий для совокупностей, из которых взяты выборки, расчет более сложен
- Проверить равенство дисперсий можно по критерию Левина (Levene's Test) : $H_0: \sigma_1 = \sigma_2$ против $H_a: \sigma_1 \neq \sigma_2$

Статистика группы

	Vehicle type	N	Среднее	Среднекв. отклонение	Среднекв. ошибка среднего
Curb weight	Automobile	115	3,18	,47	,044
	Truck	40	3,94	,69	,110



Критерий для независимых выборок

Curb weight

	Критерий равенства дисперсий Ливиня		t-критерий для равенства средних						
	F	Значимость	т	ст.св.	Знач. (двухсторонняя)	Средняя разность	Среднеквадратичная ошибка разности	95% доверительный интервал для разности: Нижняя	Верхняя
Предполагаются равные дисперсии	4,160	,043	-7,585	153	,000	-,75	,10	-,95	-,56
Не предполагаются равные дисперсии			-6,344	52,248	,000	-,75	,12	-,99	-,51